WHAT IS CLAIMED IS:

1   1.   A method for grouping log file entries by session, comprising:

2   storing a log file of entries in a memory, each of said entries identifying a client request

3   to a server;

4   retrieving a subset of log file entries from the memory;

5   identifying each entry in the memory to identify entries in the subset of log file entries

6   that belong to a complete client session;

7   grouping entries in the subset that belong to a complete client session.

1   2.   The method of claim 1, wherein a complete client session is identified by

2   identifying all entries in the subset that are associated with a particular client session and that

3   include both a beginning entry and an end entry.

1   3.   The method of claim 2, wherein an end entry is identified as any entry that

2   corresponds to a logout request.

1   4.   The method of claim 2, wherein an end entry for a client session is identified as

2   any entry associated with that client session that has no other entries for that client session that

3   occur within a session expiration window.

1   5.   The method of claim 2, wherein an end entry for a client session is identified as

2   any entry having a first timestamp value, where the difference between first timestamp value and

3   a second timestamp value associated with a subsequent entry in the subset of log files exceeds a

4   timeout value.

1   6.   The method of claim 1, further comprising outputting all entries in the subset of

2   log file entries that do not belong to a complete client session as raw log data.

1   7.   The method of claim 1, further comprising outputting as raw log data all entries in

2   the subset of log file entries that belong to an incomplete client session which has a beginning

3   entry but no end entry.

1   8.   An article of manufacture having at least one recordable medium having stored

2   thereon executable instructions and data which, when executed by at least one processing device,

3   cause the at least one processing device to:

4   read a plurality of records from a file system into a ring buffer, where said plurality or

5   records comprises a subset of all records in the file system;

6   scan each record in the ring buffer to identify a user session for said record and to

7   identify any start or end records in the ring buffer;

8   allocate, for each identified user session, an index to identify all records in the ring buffer

9   that are associated with the identified user session and to identify all start or end records; and

10   process the index to group all records in the ring buffer belonging to a complete user

11   session, to output the grouped records for further analysis.

1   9.   The article of manufacture of claim 8, wherein the index comprises:

2   a session record for each identified user session for keying into the ring buffer to identify

3   log records associated with said identified user session;

4   a hash table for keying into the session record based upon session key information;

5   a linked listing of last seen log records for each session; and

6   a linked list of first seen log records for each session.

1   10.   The article of manufacture of claim 8, wherein the ring buffer implements a

2   sliding window to process all of the log records in the file system into complete user sessions by

3   sequentially adding and removing log records to the ring buffer until all of the log records in the

4   file system have been processed.

1   11.   A system for session-based processing of log files using a data processing system

2   and network session data collected from one or more users, the system comprising:

3   a log file collection system for collecting a plurality of server request entries, wherein a

4   server request entry comprises a session identifier; and

5   a processing engine to process at least a subset of the plurality of server request entries to

6   group the server request entries by session using the session identifier in each

7   server request entry.

1   12.   The system of claim 11, wherein the processing engine uses a plurality of data

2 structures to group the web server request entries by session, said plurality of data structures

3 comprising:

4   a ring buffer for storing the subset of the plurality of web server request entries,

5   a per-session record for keying into the ring buffer

6   a hash table for keying into the per-session records

7   a linked list of last processed web server request entries for each session, and

8   a linked list of first processed web server request entries for each session.

1   13.   The system of claim 11, wherein the processing engine uses a sliding memory

2 window to process the subset of the plurality of web server request entries.

1   14.   The system of claim 11, further comprising a parser for further analysis the web

2 server request entries that have been grouped by session to generate a user session history.

1   15.   The system of claim 11, where the processing engine generates an output file

2 containing web server request entries corresponding to one or more complete user sessions.

1   16.   The system of claim 11, where the processing engine generates an output file

2 containing web server request entries corresponding to one or more incomplete user sessions.

1   17.   The system of claim 11, where the processing engine generates an output file

2 containing web server request entries corresponding to one or more user sessions that do not

3 include an end session entry.

1   18.   A system for parsing web site logs one session at a time, comprising:

2   means for storing network session data from at least one server log file;

3   means for reading a subset of the network session data;

4   means for processing the subset of the network session data to group said network session

5   data by session;

6   means for generating a first output file containing network session data grouped by

7   session; and

8   means for parsing said first output file.

1   19.   The system of claim 18, wherein the means for reading a subset of the network

2   session data comprises a sliding window.

1   20.   The system of claim 18, wherein the means for reading a subset of the network

2   session data comprises a ring buffer.